PPI Node Report

R.J. Walker, T.A. King, S.P. Joy, J.N. Mafi

MC Face-to-Face

Tucson AZ

10-11 April, 2014

PPI Node Report: Overview

- Mission Status and Highlights
- PDS4 Migration
- PPI Archive
- Next 5 Years: 2015-2019

Mission Status: Overview

Active

- Pipeline: Cassini, LRO, ***MESSENGER***, MEX
- Peer Review: MSL
- Cruise: Juno, ***MAVEN***

Planning

- Preliminary: InSight

Restoration

- Pipeline: MGS, ***Voyager***
- Design: ***ARTEMIS***

Support

- Peer Review: Dawn, New Horizons, Rosetta

Mission Status: Active Missions

MAVEN

- SIS status: 4 signed/sig. cycle; 3 draft; 3 delayed
- Preliminary peer reviews – May-Aug 2014 (MOI: Sep 2014)
- Release #1 – due 2015-Mar-02 (release: 2015-May-01)
- Raw data will form part of regular PF archive
  - Formats will be similar to calibrated products (CDF, tables)
  - The Project will archive raw telemetry at the end of Prime mission

MESSENGER

- Delivery 11 released on schedule – 7 Mar 2014
- MAG RDR, EPPS DDR certified; included in Delivery 11
- Mission extended through March 2015
  - Last archive delivery – March 2016
- Concern: Coordinate system policy

Concern: Coordinate System Policy

- MESSENGER wants to update the Mercury PCK
  - planned: body radii update
  - no other changes proposed at this time
- MESSENGER will need to begin updating data processing pipeline by Summer 2014
  - IAU can't adopt changes on this time frame
- SBN presented a whitepaper to PDS MC proposing guidelines for active mission coordinate system updates
  - Can we assume that the PDS will accept updates that comply with this proposal?

Other Activities

Dawn

- Archive Status
  - VIR
    - L-1A & L-1B Vesta: a few remaining liens
    - L-1A & L-1B Cruise: data delivered to DSC @ UCLA
  - FC
    - L-1A & L-1B Vesta: volumes certified; liens resolved
    - FC1/FC2 Calibration volumes delivered to SBN @ PSI
      - ready for cruise delta review (pending calibration volumes validation)
  - Gravity
    - L-0 data delivered to DSC @ UCLA; need to generate volumes
  - GRaND
    - L-1A & L-2 Vesta: lien-resolved documents delivered by GRaND to DSC @ UCLA
- Concern: Coordinate system for high level products

Mission Status: Restorations

ARTEMIS

- Archive will support lunar data analysis
  - 2 spacecraft: 5 instruments, 2 pseudo-instruments
  - *94,861 products (675.4 GB); 88* collections; *20* bundles (level-1 and level-2 data)
- ARTEMIS data are in CDF format
- Data will be archived under IM version 1.2.0.0
- Archive generation performed as an "active" restoration

PPI PDS4 Migration: Status

- 38 data sets partially migrated into 197 collections, 57 bundles
  - Manual migration of one-off, hand-generated products
  - Data products for large data sets to be produced with PPI tools
  - Migration performed with IM versions 1.0 and 1.1
- Final PPI migration will use IM version 1.2.0.0
  - Previously migrated products will be converted to IM version 1.2
- Developing PPI PDS4 archive design policy document
  - Defines policies and practices governing PPI PDS4 archives
  - Applies lessons learned from bundle generation exercises, MAVEN and ARTEMIS archive designs, initial migration efforts

PPI Archive: Holdings

PPI Data Holdings

	Data Sets	Data Volume	
		Current	Projected[2]
PPI Data	450	9.2 TB	10.7 TB
Non-PPI Data[1]	72	3.6 TB	3.8 TB
TOTAL	522	12.8 TB	14.5 TB

[1] "Non-PPI Data Sets" include data sets which were archived by some other entity (i.e. another PDS node, or the PSA), but which are available through the PPI web site.
[2] Through October 2014.

PPI Archive: Deliveries

PPI Data Deliveries FY2014 To-Date

Type		Quantity
Electronic Transfers	Files	14,653,300
	Volume	442.6 GB
Media Transfers	Deliveries	3
	Volume	639.2 GB
Special Processing	Requests	0
	Hours	0

PPI Archive: Preservation

PPI Local Archive Copies

Location	Type	Holdings	Backup	Storage
PPI Node (UCLA)	Primary	Full (522 Data Sets, 12.8 TB)	N/A	RAID 6
PPI Node (UCLA)	Mirror	Full (522 Data Sets, 12.8 TB)	Sync with primary	RAID 6
PPI Sub-node (U. Iowa)	Online Backup[1] (compressed)	Full (522 Data Sets, 12.8 TB)	Mar 2014	RAID 5
Local offsite copy	Offline Backup (compressed)	Full (522 Data Sets, 12.8 TB)	Mar 2014	Hard Drive

[1] Transitioning to RAID 6 system which will function as a mirror and will sync with primary.

PPI Archive Maintenance

- Mimic tools developed and deployed
  - File synchronization tool
  - Designed for large scale archives (Big Data)
    - Faster than rsync, scp, bbcp, git
  - Supports multiple online copies at multiple locations
    - Supports push or pull mirroring
    - Secure transfer (uses ssh with certificates)
  - Includes report generation
  - We can verify our entire archive in 11 minutes

PPI Archive: Preservation (cont.)

PPI Deep Archive copies

Location	Type	Holdings
NSSDC (GSFC)	Electronic	Partial (363 Data Sets, 10.5 TB)
NSSDC (GSFC)	Hard Media	Partial (291 Volumes, 0.3 TB)

Details

In preparation: 44

Ready for submission: 40 (13 in process)

Accepted by NSSDC: 363

PPI Software and Services

- New PDS/PPI webpages deployed (Dec. 2013)
  - PDS 3.8 metadata
  - Data holdings organized by data set
  - PDS labeled table browsing
  - Format conversion for PDS labeled tables to
    - VOTable
    - CSV
- Data Provider Tools
  - igpp.docgen – label generation
    - Velocity templates with added support for common utility classes and data from PDS3 labels, CSV, keyword/value and CDF inputs.
  - pds.cdf – extract metadata, verify PDS compliance

Security

- No new incidents
- Disaster Recovery Plans/Updates
  - Redundancy – multiple coordinated servers
  - Replication – on-line on-site mirror, on-line off-site mirror and off-line copy
  - Pro-active monitoring of hardware and archive
  - Regular review
- IT security planning
  - Multiple, multi-layered firewall
  - Proxy access to services and archive
  - Log monitoring

Staff Time Commitments

PDS4 Activity Support (Current)

Person	Design		Implementation	
	CCB	DDWG	Tool Dev.	Product Dev.
T. King	2%	10%	20%	5%
S. Joy	1%	1%	-	5%
J. Mafi	-	20%	5%	20%
D. Kazden	-	20%	5%	50%

- Past DDWG effort shifting from design to implementation
- Product Development – includes mission PDS4 support, migration work, and node policy and procedure planning

PPI Plans: 2015-2019

Projected Archive Growth

PPI Archive Totals	2015	2016	2017	2018	2019
Active Missions	11	11	10	10	8
Total Data Sets	75	80	85	90	95
Archive Volume (TB)	16.5	18.5	19.5	20.5	21.5

PPI Plans: 2015-2019

Mission Support/Data Acquisition

- Support active and future missions
- Explore inclusion of simulation data in the archive
- Continue restoration efforts (ARTEMIS, Voyager, etc.)

PDS4 Migration

- Complete migration of full PPI holdings

Tool and Services

- Continue transition to the new PDS4 paradigm

PPI Plans: 2015-2019

Website

- Implement Mission Pages for all PPI holdings
- Visualization
  - Continue working with AMDA (CDPP), TOPCAT (CNES), and Autoplot (U. Iowa) groups to develop access API's to enable on-the-fly visualization of PPI data using those tools.
  - Deploy UCLA "Web Splash" visualization tool
- Format transformation
  - CDF to ASCII/PDS labeled to CDF
- Coordinate system transformation
- Improved search capability
- PDS4 service and infrastructure integration

BACKUP SLIDES

Mission Status: Active Missions

Juno

- 3 of 4 F&P SIS's signed (MAG missing)
- First archive release (Earth flyby) – April 2015

MSL

- RAD EDR & RDR data sets certified, RDR still working liens; in production

Cassini

- CAPS – ion & electron moments, s/c potential data sets delivered, online (uncertified)
- MAG – Users' guide accepted; online; needs revision

Mission Status: Restorations

Galileo

- PLS – Ion moments; MOM_INPUT files in preparation (under DAP funding)

Voyager

- Acquisition of cruise data continues

Mission Status: Support

New Horizons

- Supported SWAP & PEPSSI post-Jupiter cruise peer review
- May need to support 2011-2013 peer review later this year

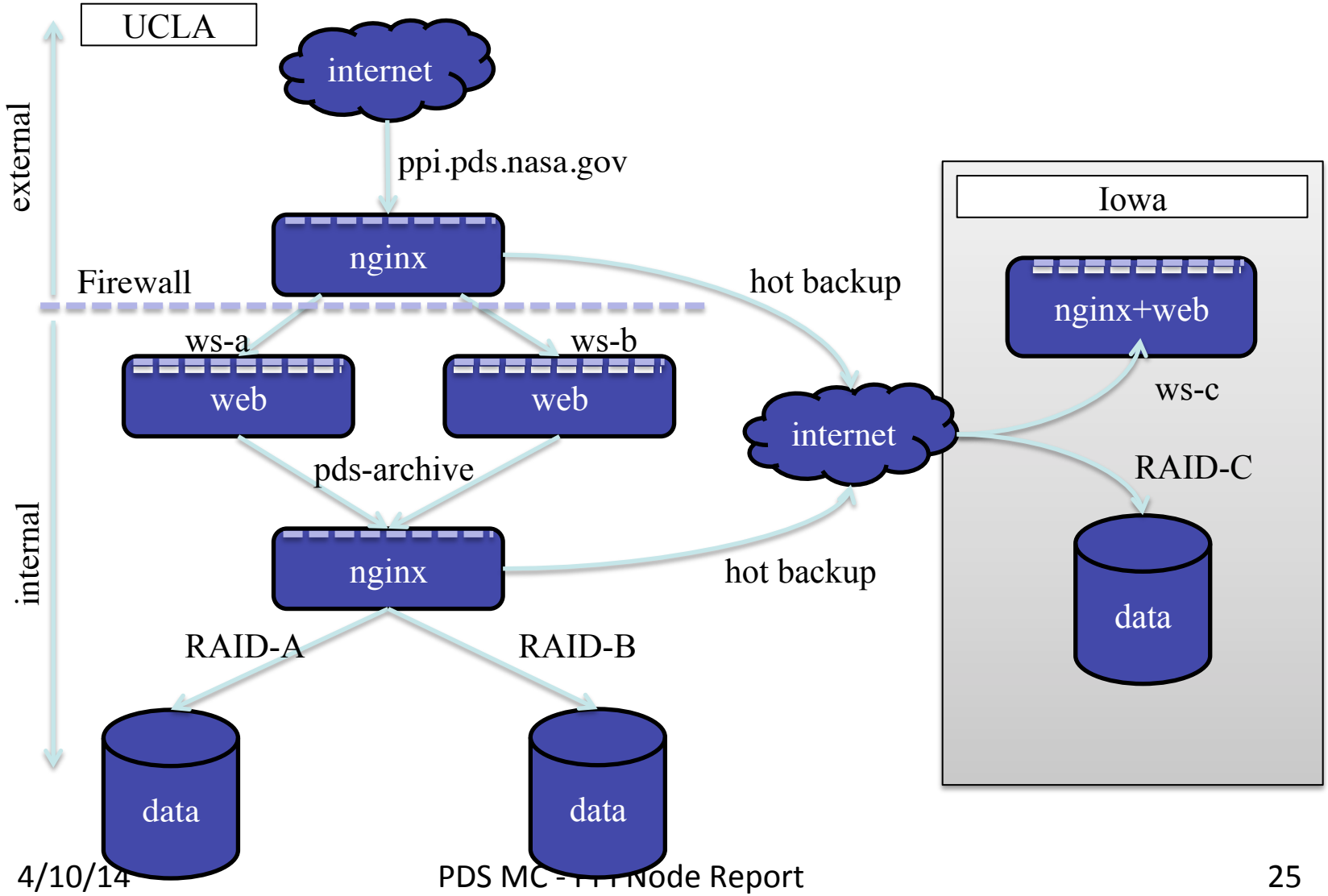
Rosetta

- Arrives at comet in 2014; may need to support peer review later 2014 or early 2015

PDS4 Archive CDF Requirements

1) CDF version 3.4 or later.
2) MSB (Network) encoding for all values.
3) CDF Tool compliant metadata.
4) ISPT/IACG compliant metadata.
5) Use single file CDF.
6) No compression (file or variable).
7) No fragmented variables.
8) No sparse variables. All data values are physical.
9) No unused records.
10) Use only “zVariables”
11) No virtual (calculated) variables.

PPI Archive Architecture



MAVEN Archive Schedule

Date	Milestone
2014-Mar-24	Signed SIS
2014-Apr-18	Sample data products due
2014-May through 2014-Aug	Preliminary PDS peer review (SIS and sample data products)
2014-Nov-01	Start of Mars Science Ops
2015-Mar-02	Release #1: Data due to PDS
2015-Mar through 2015-Apr	Release #1: PDS peer review
2015-May-01	Release #1: Public release

PDS4 Tools Descriptions (1/3)

Tool	Description	Developer
AMDA	Web space physics visualization tool	CDPP
Autoplot	Web visualization tool for numerous data types (including VO Table)	Cottage
CDF to ASCII table/ PDS labeled to CDF	Transforms PDS labeled CDF to PDS labeled ASCII tables and viseversa	UCLA
cdfconvert	Converts a CDF file's CDF library version compliance and format	GSFC
cdfdump	Dumps the data and/or metadata of a CDF file in a readable form	GSFC
ditdos/write	Performs data format transformations: PDS labeled table to VO Table	UCLA
Generate Tool	Generates PDS4 labels from either a PDS3 label or PDS3-specific Document Object Model (DOM) object	JPL

PDS4 Tools Descriptions (2/3)

Tool	Description	Developer
Harvest Tool	Collects metadata from PDS4 product labels for the Registry Service and Search Tool	JPL
igpp.docgen	Populates an Apache Velocity template using metadata scanned from an PDS label, XML, or CDF file producing a PDS label, XML, or plain text file	UCLA
oXygen, etc.	XML editor and debugger	COTS
pds.cdf.CDF	Reads and returns CDF metadata	UCLA
pds.cdf.Check	Reads a CDF file and verifies conformance with the PDS archive requirements for CDF	UCLA
Registry Service	Software for tracking, auditing, locating, and maintaining artifacts within the system	JPL
PDS4 Tools	PDS4 object access, and transformation tool	ARC

PDS4 Tools Descriptions (3/3)

Tool	Description	Developer
Search Tool	Performs product searches based upon databases created by the Harvest Tool	JPL
Table display	Reads and displays a PDS4 labeled table	ARC
TOPCAT	Web visualization tool for numerous data types (including FITS, VO Table, and CDF)	CNES
Transform Tool	Transforms PDS3 and PDS4 product labels and product data into common formats	JPL
Validate Tool	Validates PDS4 product labels and product data	JPL
Web SPLASH	Web visualization tool for PDS labeled data	UCLA
XML generation/ transformation tool	Generates a PDS4 label from	NMSU

Mission Pages